Auditory and visual capture during focused visual attention

Koelewijn, T.; Bronkhorst, A.W.; Theeuwes, J.

published in
Journal of Experimental Psychology: Human Perception and Performance
2009

DOI (link to publisher)
10.1037/a0013901

document version
Publisher's PDF, also known as Version of record

Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 5, 1303–1315

0096-1523/09/$12.00 DOI: 10.1037/a0013901

Auditory and Visual Capture During Focused Visual Attention

Thomas Koelewijn
Vrije Universiteit Amsterdam

Adelbert Bronkhorst
Vrije Universiteit Amsterdam and TNO Human Factors

Jan Theeuwes
Vrije Universiteit Amsterdam

It is well known that auditory and visual onsets presented at a particular location can capture a person's visual attention. However, the question of whether such attentional capture disappears when attention is focused endogenously beforehand has not yet been answered. Moreover, previous studies have not differentiated between capture by onsets presented at a nontarget (invalid) location and possible performance benefits occurring when the target location is (validly) cued. In this study, the authors modulated the degree of attentional focus by presenting endogenous cues with varying reliability and by displaying placeholders indicating the precise areas where the target stimuli could occur. By using not only valid and invalid exogenous cues but also neutral cues that provide temporal but no spatial information, they found performance benefits as well as costs when attention is not strongly focused. The benefits disappear when the attentional focus is increased. These results indicate that there is bottom-up capture of visual attention by irrelevant auditory and visual stimuli that cannot be suppressed by top-down attentional control.

Keywords: cross-modal, attentional capture, enhanced processing, exogenous, endogenous

Covert visual attention can be directed to a specific location in the visual world without making eye movements. This direction can happen voluntarily, by steering attention endogenously to that location, or automatically, when attention is exogenously captured. Endogenous attention has been compared metaphorically to a spotlight that casts its light on relevant visual information (Broadbent, 1982; Posner, Snyder, & Davidson, 1980) and that can be directed to a target location by presenting an informative cue prior to the target. An example of such a cue is an arrow displayed at the center of the visual field that points with a high probability (e.g., 80% valid) to a possible target location. Exogenous capturing of attention can for example be evoked by a visual onset occurring at the target location. Exogenous cues can shorten reaction times to targets even when the cues do not reliably predict (with a validity at chance level) the location of the upcoming target (e.g., Jonides, 1981; Yantis & Jonides, 1984). Exogenous cuing is not restricted to the visual modality alone: Tactile (Posner, 1978; Spence & McGlone, 2001) and auditory cuing effects (Spence & Driver, 1994) have also been reported. Both endogenous and exogenous cues can cause an object appearing at an attended location to be detected faster and more accurately than an object appearing at an unattended location (Posner, 1980; Posner et al., 1980). However, peripheral exogenous cues capture attention automatically, whereas central endogenous cues seem to be less obligatory (e.g., Jonides, 1981; Theeuwes, 1991; Yantis & Jonides, 1984).

Both endogenous and exogenous cuing effects can be described by the allocation of attentional resources to a cued location. For endogenous cuing, these attentional shifts are controlled in top-down fashion, and for exogenous cuing, these shifts are enforced by bottom-up processes. Although these processes of endogenous and exogenous visual attention have mostly been studied in separate paradigms, there have also been studies that looked at the interactions between both attentional processes (Jonides, 1981; Muller & Rabbitt, 1989; Theeuwes, 1991; Yantis & Jonides, 1990). Theeuwes (1991) investigated the relation between exogenous and endogenous visual attention within a single paradigm. In this study, participants had to identify a target letter among three distractor letters all positioned equidistantly on an imaginary circle. An endogenous cue (a centrally presented arrow) reliably indicated the location of the target. In addition, a nonpredictable exogenous visual onset cue was presented near one of the letters. When the central arrowhead was presented after the exogenous cue, attention was drawn to the location of the exogenous cue. But when the central arrowhead was presented prior to the presentation of the exogenous cue, attention was in a focused state and therefore the exogenous cue had no effect. These results show that exogenous capture of attention can cease to exist when attention is endogenously focused on a location in space. In other words, typically, exogenous events do not cause interference when presented outside the focus of attention (for similar results, see Yantis

Thomas Koelewijn, Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; Adelbert Bronkhorst, Cognitive Psychology, Vrije Universiteit Amsterdam, and TNO Human Factors, Soesterberg, the Netherlands; Jan Theeuwes, Cognitive Psychology, Vrije Universiteit Amsterdam.

This research was supported by Dutch Technology Foundation STW Grant VET.7079, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs, to Jan Theeuwes and Adelbert Bronkhorst.

Correspondence concerning this article should be addressed to Thomas Koelewijn, Cognitive Psychology, Vrije Universiteit, van der Boechorststraat 1, 1081BT Amsterdam, the Netherlands. E-mail: t.koelewijn@psy.vu.nl

& Jonides, 1990). The observed suppression of visual exogenous events suggests that visual exogenous attention is not a completely automatic process.

Endogenous and exogenous cuing effects across modalities were demonstrated in two studies by Spence and Driver (1996, 1997). In these studies, participants had to perform an orthogonal cuing task in which they had to make an elevation judgment (up vs. down) regarding auditory or visual targets presented to the left or right of fixation. Loudspeakers and LEDs mounted in front of the loudspeakers were used to generate target stimuli. In the first study (Spence & Driver, 1996), participants were presented with an endogenous cue (a central arrowhead) indicating that the target was more likely to appear on the side indicated by the cue. When the cue was valid, participants were faster to make an elevation judgment to that side, regardless of the modality in which the target was presented. In the second study (Spence & Driver, 1997), the target side was cued (exogenously) at chance level by a visual (LED onset) or auditory (pure tone) cue presented at the same eccentricity as the targets. For the cross-modal conditions (visual cue and auditory target or auditory cue and visual target), there was only a cuing effect when a visual target was preceded by an auditory cue but not vice versa. For the unimodal conditions (visual cue and visual target or auditory cue and auditory target), both visual and auditory cuing effects were found. Spence, McDonald, and Driver (2004) attributed this observed asymmetry to a higher spatial resolution of the visual compared with the auditory perceptual system. Presumably, this difference in resolution is associated with a corresponding difference in the size of the spatial area that is attended. According to Spence and colleagues (2004), "when testing for visual-upon-auditory effects, the auditory targets were in effect presented too far away in external space (in elevation) from the preceding visual cue for any crossmodal cuing effect to have been observed" (p. 286). In other cross-modal cuing studies, McDonald and colleagues (McDonald & Ward, 2000; Ward, McDonald, & Lin, 2000) used a go/no-go task in which cues and targets were presented at the same spatial location. Because they now found a cuing effect for auditory targets preceded by visual cues, their results are consistent with the above explanation of Spence et al. for the earlier observed asymmetry in cross-modal cuing.

The finding of cuing effects across modalities raises the question of whether unimodal and cross-modal cuing effects are based on similar processes. One way to investigate this issue is to test whether effects reported in unimodal conditions would also apply to cross-modal conditions. For example, if cross-modal attention is similar to unimodal attention, one would expect similar effects as reported by Theeuwes (1991). In other words, both auditory and visual exogenous events should cease to capture attention when visual attention is in a focused state.

In a recent study, van der Lubbe and Postma (2005) tested this notion. They used a variation on the orthogonal cuing task. Participants had to perform a discrimination task on targets presented as arrowheads pointing up or down. These targets were displayed on LED grids placed on the left and right side at 28.3° or 19.3° visual angle relative to a centrally positioned LED grid used for fixation. One second prior to the onset of the target, the central grid displayed an arrow pointing to the correct target location (100% valid) or displayed a noninformative (neutral) cue. An exogenous visual or auditory cue was presented 200 ms before the target, indicating the target location at chance level. Unlike Theeuwes (1991), van der Lubbe and Postma (2005) found that abrupt onsets of both visual and auditory cues captured attention even in conditions in which observers were focused on the endogenously cued location. Note, however, that van der Lubbe and Postma (2005) used quite large visual angles between fixation and targets (28.3° or 19.3°), whereas the angle in the Theeuwes (1991) experiment was only 4.2°. Even though van der Lubbe and Postma (2005) failed to replicate Theeuwes's (1991) original results, they found results for unimodal and cross-modal cuing that are comparable to each other. Therefore, their results are consistent with the idea that unimodal cuing and cross-modal cuing basically operate according to the same underlying mechanisms.

In a recent study by Mazza, Turatto, Rossi, and Umilta (2007), participants performed an orthogonal cross-modal cuing task similar to the one used by Spence and Driver (1997). In their first experiment, the target side was random and results (for a 150-ms cue–target interval) showed both unimodal visual and auditory cuing effects, but only a cross-modal cuing effect when a visual target was preceded by and auditory cue. These results replicate the cross-modal cuing asymmetry shown by Spence and Driver (1997). In their second experiment, the target side remained the same during an entire block. Thus, participants knew at which location the target would appear although they were also presented with nonpredictive exogenous cues. The results (for the 150-ms cue–target interval) show a cross-modal cuing effect for visual targets preceded by valid (512 ms) or invalid (530 ms) auditory cues. Surprisingly, an opposite effect was found when auditory targets were preceded by valid (629 ms) or invalid (586 ms) visual cues. Note that no unimodal cuing effects were observed, which is in contrast to the results of their first experiment. Mazza and colleagues (2007) therefore did not find a unimodal cuing effect when attention was in a focused state, basically replicating Theeuwes's (1991) results. In addition, they showed that when a visual target is preceded by an auditory cue, the cross-modal cuing effect holds even during visual focused attention. Thus, although these results differ from those of van der Lubbe and Postma (2005) for the unimodal conditions, both studies agree that exogenous cross-modal cuing cannot be suppressed by focused attention. However, it is not yet clear why results for cross-modal cuing should deviate from the classic visual exogenous cuing effects demonstrated by Theeuwes (1991) and Yantis and Jonides (1990).

A recent study by Santangelo and Spence (2007) used an orthogonal cuing paradigm similar to the one used by Spence and Driver (1997). Their design, however, contained only elevation judgments concerning visual targets. In addition to the elevation judgment task, a second task was introduced to keep the participants' attention focused to the center of the display. In this additional task, observers had to respond to digits embedded in a stream of letters presented in a rapid serial visual presentation (RSVP). In the high-load condition, a target digit was presented centrally in 67% of the trials and a peripheral target for the elevation judgment task was presented in the remaining 33% of the trials. In the no-load condition, no RSVP stream was presented. In all trials, a peripheral visual, auditory, or bimodal (visual and auditory) exogenous cue was presented on the left or the right side. With respect to the position of the peripheral targets, these cues could be either valid or invalid. The results for the no-stream condition showed auditory, visual, and bimodal cuing effects.

However, in the high-load condition, only a bimodal cuing effect was observed. These results indicate that unimodal and cross-modal exogenous cuing effects disappear when, in addition to the cuing task, participants focus their attention on an RSVP stream containing a possible target. Santangelo and Spence (2007) conducted an additional experiment confirming that disengagement from the RSVP stream could be ascribed to multisensory integration instead of the double perceptual input provided by the bimodal cue. In this experiment, redundant visual and auditory cues were compared with the bimodal cue. Again, these outcomes showed only a cuing effect for bimodal cues. The authors concluded that unimodal but not bimodal exogenous cuing effects can be suppressed by means of endogenous focused attention.

In summary, van der Lubbe and Postma (2005) showed that there are both cross-modal and unimodal exogenous cuing effects when attention is focused endogenously, in contrast to Theeuwes (1991), who found no unimodal cuing effects, and Santangelo and Spence (2007), who found suppression of both unimodal and cross-modal exogenous cuing. Mazza and colleagues (2007) showed suppression of unimodal but not cross-modal cuing. It is not clear to what degree these results can be explained by methodological differences. Important factors may be where and how strongly attention was focused endogenously, but these factors do not seem to correlate well with the results. For example, Santangelo and Spence (2007) not only required participants to focus their attention close to fixation (whereas all other studies used peripheral locations) but also did this in a way that presumably caused stronger attentional focus. Although this might explain why their results for cross-modal cuing deviate from those of the others, it is then difficult to understand why their results for within-modality cuing are in essence the same as those of Theeuwes (1991) and Mazza and colleagues (2007). Another issue is that all studies quantify attentional capture by taking the difference between response times for valid and invalid exogenous cues. It is typically assumed that effects of endogenous and exogenous cuing are due to attention shifts to or away from the target (Jonides, 1981; Posner, 1980; Spence & Driver, 1996). However, it is not clear whether this holds for the observed cuing effects found in these and other studies (e.g., McDonald & Ward, 2000; Spence & Driver, 1997). More specifically, one can question whether cross-modal cuing effects are the result of costs, by attention being drawn away (captured) to an invalid cue location, and/or of benefits, by attention being captured by a valid cue location. No previous attempts have been made to separate these two components, although it seems evident that a factor such as the strength of the (endogenous) attentional focus will affect them differently. A further issue that complicates comparison of the above studies is that eye movements were not always registered (e.g., Mazza et al., 2007). In particular when participants are instructed to endogenously focus their attention on a location prior to the presentation of a target, one cannot exclude the possibility that they will make eye movements (thus introducing a confounding factor).

In order to shed more light on these issues, we have conducted a series of experiments based on the classic orthogonal cuing paradigm introduced by Spence and Driver (1994). We used both exogenous and endogenous cues, and we modulated the degree to which observers focused their attention on the target location. This modulation was achieved by changing the validity of the endogenous cue and by using placeholders that indicate the precise spatial region where the target is about to appear. Furthermore, we included a baseline condition in which the exogenous cue provided temporal but no (reliable) spatial information (for elaborate explanation on this topic see, Jonides & Mack, 1984). This condition enables us to separate performance costs due to invalid cues from benefits caused by valid cues. A spatially neutral auditory cue was created by simultaneously presenting two uncorrelated noise bursts from two loudspeakers, which in our set-up were located to the left and right of the monitor on which the visual stimuli were presented. This set-up causes a broad spatial percept in front of the participant, which extends to the sides beyond both loudspeakers (Blauert, 1997). Note that the use of uncorrelated signals is essential here because two correlated signals are perceived as an easily localizable sound in the middle between the two loudspeaker positions (in our case this would be directly in front of the participant) as a result of summing localization (Blauert, 1997). This principle for creating spatially neutral auditory cues was used in all our experiments. In our final experiment, we also presented a neutral visual cue that was created by simultaneously displaying cues on both possible target locations. In all experiments except the first one (which did not include endogenous cues), eye movements were monitored to make sure all observed cuing effects could be attributed to covert attention.

In our first experiment, we validated our paradigm and specific set-up by replicating the cross-modal cuing effects found earlier by Spence and Driver (1997) and others. Additionally, we introduced the spatially neutral auditory cue that allowed us to specify the observed cuing effect in terms of costs and benefits. In Experiments 2 and 3, endogenous cuing was added by means of a central arrowhead presented before the auditory cue indicating the correct target side in 80% or 100% of the trials, respectively. In Experiment 4, we investigated whether the addition of placeholders, which are assumed to induce an even stronger attentional focus on the position of the target side before its appearance, would alter the previously observed cuing effect. Finally, in Experiment 5, we directly compared effects of cross-modal (auditory) and unimodal (visual) cuing in conditions with and without endogenous focus of visual attention.

It should be noted that there is an ongoing discussion concerning the use of arrowheads as endogenous cues (Hommel, Pratt, Colzato, & Godijn, 2001; Santangelo & Spence, 2008). We are aware that some studies show that arrowheads can also have an exogenous cuing effect when presented at chance level. This effect is probably because overlearned symbols are almost automatically processed (Hommel et al., 2001) and are therefore directing attention partly in a bottom-up fashion. Possibly, there are also exogenous effects of the arrowheads that we have used, but because they were presented at least 650 ms before the target and were always followed by an exogenous location cue, it is unlikely that they have influenced our results.

Experiment 1

The task in this experiment was similar to the orthogonal cuing task used by Spence and Driver (1997). Instead of using LEDs, visual stimuli were presented on a computer screen. The loudspeakers that generated the auditory cues were located to the left and right of the computer screen (for a comparable set-up, see Mondor & Amirault, 1998; Talsma & Woldorff, 2005b). We used only auditory cues in combination with visual targets and presented the cues at a stimulus onset asynchrony of 200 ms, because this condition yielded a large cuing effect in earlier studies (e.g., Spence & Driver, 1997). In our paradigm, we also used an auditory spatially diffuse cue, which served as a neutral condition; that is, it did not seem to emanate from a specific direction.

Method

Participants. Twelve students of the Vrije Universiteit Amsterdam (6 men, 6 women; mean age = 21.4 years, range = 18–28 years) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed about the experimental procedure and were naive to the purpose of the experiment.

Apparatus and design. Participants were seated in a dimly lit room at approximately 80 cm from a computer screen (CRT, 17 in., 120 Hz). The experiment was run in E-Prime 1.1 (SP3) (Psychology Software Tools, 1996-2003). To the left and to the right of the screen, a loudspeaker was placed at an angle of 18.3° from fixation, and both loudspeakers were aligned to the vertical middle of the screen. The experiment consisted of five blocks containing 36 trials each. There was a valid condition (33% of the trials) in which the auditory cue and visual target were presented on the same side, an invalid condition (33%) in which the auditory cue and visual target were presented on opposite sides, and a neutral condition (33%) in which the auditory cue could not be assigned to a specific location in space (but still provided the same temporal information as the other cues). All conditions were presented randomly within blocks; the first block was for practice purposes, leaving 48 trials for each condition.

Procedure and stimuli. Figure 1 gives an example of a typical trial. At the beginning of each trial, a white fixation dot (diameter 0.2°) appeared and stayed on screen until a response was made. Participants were instructed to fixate on this dot during the entire trial and to refrain from making eye movements. After a random delay time of 400 ms to 650 ms, an auditory cue consisting of a white noise burst was presented for 100 ms. This cue came equiprobably from the left or right loudspeaker and was valid or invalid with respect to the target location, or it was neutral and came from both loudspeakers at the same time. In the latter case, two uncorrelated noise bursts were used. The cues presented through a single loudspeaker were boosted by an extra 3 dB, to create the same subjective loudness as the neutral cue that was presented through two loudspeakers. Two hundred milliseconds after the onset of the auditory cue, a visual target consisting of a white dot (diameter 0.2°) was presented for 140 ms. There were four possible target locations: two locations were positioned 10.5° to the left of fixation, and two locations were positioned 10.5° to the right of fixation. The two locations at each side were positioned above each other, one 2.4° above and the other one 2.4° below the vertical center of the screen. The target appeared at one of these four locations at chance level. The participants' task was to report in a speeded but accurate fashion whether the target appeared above or below the vertical centre of the screen by pressing the number *8* or number *2* key, on the number pad of a *QWERTY* keyboard, respectively. Participants responded with both index fingers and were free to choose which finger to use for which button as long as they kept it the same during the experiment.

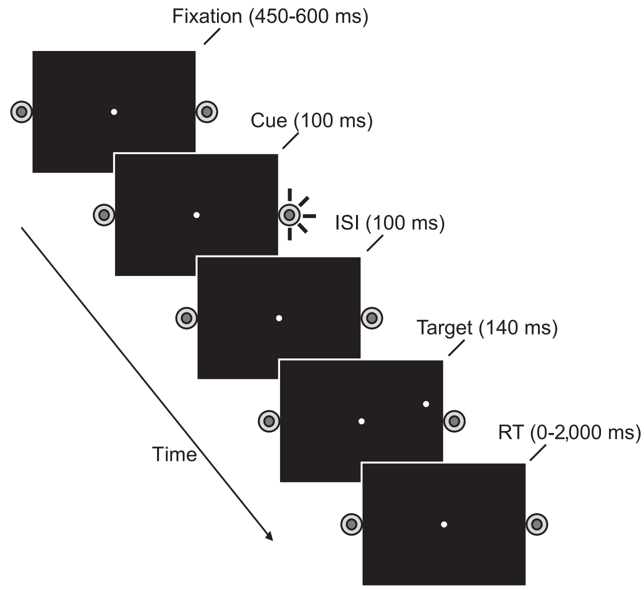

Figure 1. Schematic representation of the paradigm used. Participants performed an orthogonal cuing task in which they had to discriminate between targets presented above or below the vertical middle of the screen. Targets were presented on the left or right side of the screen and were preceded with a stimulus onset asynchrony of 200 ms by a nonpredictable auditory cue coming out of the left or right loudspeaker. ISI = interstimulus interval, RT = reaction time.

Because of the task's orthogonal design, there was no need to balance out for possible motor response effects. Responses had to be made within a time window of 2,000 ms after target onset. After the response, participants received feedback when they had made an error—the fixation dot turned red for 150 ms. After each trial, an interval of 850 ms followed before the next trial started. Following each block, participants received feedback in the form of a percentage correct score and a mean reaction time. At the beginning of the experiment, participants were told that the auditory cues would provide no information about the location of the targets and therefore could be ignored.

Results

Premature (<200 ms) and slow (>1,000 ms) responses (in total 1.3%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (92.6%) were calculated for each participant for each condition. Figure 2 presents the mean reaction times for each condition (valid, 343 ms; invalid, 365 ms; and neutral, 352 ms) averaged over participants. The error bars in this figure represent the .95 confidence interval (5.8 ms) for the exogenous cuing main effect, following Loftus and Masson (1994). An overview of the mean reaction times, their standard deviations, and the mean error scores, for each condition and for all experiments, is shown in Table 1. An analysis of variance (ANOVA) on reaction time with cue validity (valid, invalid, neutral) as a factor revealed a significant effect, $F(2, 22) = 15.808$, $MSE = 92.210$, $p < .001$. Three pairwise two-tailed t tests between the cuing conditions were conducted. Valid compared with invalid ($p < .001$), neutral compared with valid ($p = .006$),

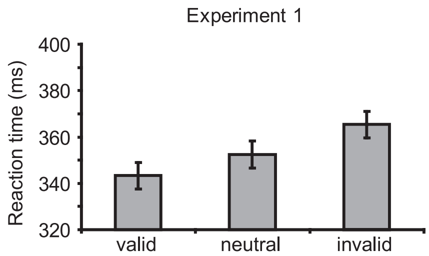

Figure 2. Results of Experiment 1 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cuing main effect (Loftus & Masson, 1994).

and neutral compared with invalid conditions (p = .021) all differed significantly.

The mean error scores (valid, 5.9%; invalid, 8.3%; and neutral, 8.1%) were also calculated. An ANOVA on error scores with cue validity (valid, invalid, neutral) as a factor showed no reliable effect.

Discussion

The current experiment shows a cuing effect similar to that observed by Spence and Driver (1997). An extension with respect to earlier studies was that it included a neutral condition, allowing separation of reaction time costs and benefits. Our results indicate that the typical costs and benefits of cross-modal cuing can be attributed to shifts of spatial attention just as in the classic way of explaining unimodal visual cuing effects (Jonides, 1981; Posner, 1980; Spence & Driver, 1996). The explanation goes as follows: First, directing attention is a fast but not an instantaneous process. In order to direct attention, it has to be disengaged from its old location, shifted to a new location, and then engaged on this new location (Posner et al., 1980). When a target location is validly cued, attention is already directed to the target location before the presentation of the target, which results in more attentional recourses at the target location, allowing for easier processing of the target. The benefit of this is reflected in a faster response relative to the neutral condition. When, on the other hand, a target location is invalidly cued, attention is directed to a location opposite to the target location, which results in fewer attentional resources being available on the target location for processing of this target and slower response times relative to the neutral condition.

The results of Experiment 1 thus show that cross-modal cuing is associated with costs and benefits that can be explained by shifts of spatial attention. As discussed in the introduction, within the visual domain these costs can be suppressed when attention is focused before the presentation of the exogenous cue (Mazza et al., 2007; Theeuwes, 1991; Yantis & Jonides, 1990). In order to test whether this same principle holds for cross-modal cuing, we conducted a second experiment in which an endogenous visual cue in the form of a centrally presented arrowhead was shown prior to the exogenous auditory cue. Because this cue was a relatively small centrally presented symbol and appeared at least 650 ms before the target, we did not expect that this cue exerted an additional exogenous effect (Jonides, 1981). One of the consequences of using an endogenous cue with a longer cue-to-target interval is that

Table 1
Mean Reaction Times (in Milliseconds), Standard Deviations, and Percentages of Errors for Experiments 1–5

Exogenous cue modality and endogenous cue condition	Exogenous cue validity: Valid *M*	Valid *SD*	Valid % error	Neutral *M*	Neutral *SD*	Neutral % error	Invalid *M*	Invalid *SD*	Invalid % error
Experiment 1									
Auditory									
No cue	343	32	5.85	352	34	8.14	365	32	8.26
Experiment 2									
Auditory									
80% valid cue	349	36	3.02	357	40	3.61	372	41	5.21
20% invalid cue	378	32	5.96	398	41	8.33	402	41	8.04
Experiment 3									
Auditory									
100% valid cue	335	31	2.32	343	36	4.71	351	43	4.27
Experiment 4									
Auditory									
100% valid cue	362	33	7.10	364	39	6.77	373	36	6.50
Experiment 5									
Auditory									
No cue	394	26	4.14	411	28	5.16	428	31	8.56
100% valid cue	392	41	4.32	395	39	5.62	407	39	5.41
Visual									
No cue	423	30	5.09	439	32	5.38	484	37	5.58
100% valid cue	410	25	5.98	422	30	6.20	467	36	4.74

it gives participants time to make eye movements toward the target. When eye movements are made, the effects are not solely produced by covert visual attention. In order to control for possible overt orienting of attention, eye movements were recorded in the following experiments.

Experiment 2: Endogenous Cue 80% Correct

As explained above, an endogenous visual cue presented prior to the exogenous auditory cue was used in this experiment. The cue was an arrowhead presented in the center of the screen that allowed the participants to focus their visual attention on the target location before the onset of the peripheral auditory cue. When an auditory cue does not automatically capture attention, no costs on invalidly exogenous cued trials are expected. However, when the auditory exogenous cue is still able to capture attention even when participants are highly focused on a location in space, costs on invalidly cued trials are expected. If the cuing effects resulting from the presentation of an auditory exogenous event are due only to shifts of spatial attention, then one expects no benefits of the auditory cue when presented at a location at which participants are already focused. Therefore, we expect no benefits on valid trials. To ensure that participants indeed used the endogenous cue, we made it valid on the majority (80%) of trials.

Method

Participants. Twelve new students from the Vrije Universiteit Amsterdam (2 men, 10 women; mean age = 20.7 years, range = 18–30 years) participated in the experiment.

Stimuli and design. The task was basically identical to the one used in Experiment 1 except that at the start of each trial a visual cue, an arrowhead appointing to the left or right (width 0.5° and height 0.4°), was presented for 600 ms. In 80% of the trials, this arrowhead indicated the side where the subsequent target would appear. The combination of two types of endogenous visual cues (valid or invalid) with three possible auditory cues (valid, invalid, and neutral) resulted in six conditions. A total of nine blocks containing 60 trials each were presented during the experiment. All conditions were presented randomly within blocks, and the first block was for practice purposes only. Thus, there were 128 trials for each exogenous cue condition (valid, invalid, and neutral) when the endogenous cue was valid (80%) and 32 trials for each exogenous cue condition when the endogenous cue was invalid (20%).

Electrooculogram recoding and analysis. The horizontal and vertical electrooculograms were recorded bipolarly by electrodes located on the outer canthi of both eyes and a pair of electrodes on the supraorbital and infraorbital ridge of the right eye, respectively. Recordings were made at a 500-Hz sampling rate. For detecting eye movements, a spike detection algorithm was used (for a full description, see Talsma & Woldorff, 2005a). In short, this algorithm uses a sliding time window (sliding with 2-ms steps) set to 100 ms in which the maximum amplitude differences are calculated between all possible time point combinations within the window. The maximum allowed amplitude difference was set to 70 μV. All trials showing larger amplitude differences, from the onset of the visual cue till the offset of the target, were excluded from analysis. An amplitude range of 70 μV filtered out trials containing eye movements larger than 5° horizontal angle (Peelen, Heslenfeld, & Theeuwes, 2004), less than half the angle needed to focus on the target location. This range prevented leaving out trials containing small muscle artifacts unrelated to eye movements. To reduce the possible loss of trials on the basis of eye blink artifacts (causing large amplitude differences sometimes hard to distinguish from eye movements), an intertrial interval of 2,000 ms instead of 850 ms was used, and participants were instructed to blink their eyes during this period.

Results

Trials containing eye movements (8.6%) and premature (<200 ms) or slow (>1,000 ms) responses (0.6%) were removed from further analysis. For the remaining trials there were on average 95.3% correct responses. For these trials we calculated per participant the mean reaction times for each condition. The mean reaction times for each condition averaged over participants are plotted in Figure 3. The error bars in this figure represent the .95 confidence interval (6.4 ms) for the exogenous cuing main effect. An ANOVA conducted on the reaction times with visual cuing (valid and invalid) and auditory cuing (valid, invalid, and neutral) as the within-subject variables showed a main effect of visual cuing, $F(1, 11) = 48.749$, $MSE = 408.216$, $p < .001$, and of auditory cuing, $F(2, 22) = 30.052$, $MSE = 115.332$, $p < .001$. In addition, the interaction between visual and auditory cuing was reliable, $F(2, 22) = 4.781$, $MSE = 69.451$, $p < .05$.

Significant differences in the visual valid condition were shown for auditory valid (349 ms) compared with auditory invalid conditions (372 ms; $p < .001$), auditory neutral (357 ms) compared with auditory valid conditions ($p = .004$), and auditory neutral compared with auditory invalid conditions ($p < .001$). For the visual invalid condition, both auditory valid (378 ms) compared with auditory invalid conditions (402 ms; $p < .001$) and auditory neutral (398 ms) compared with auditory valid conditions ($p = .003$) showed a

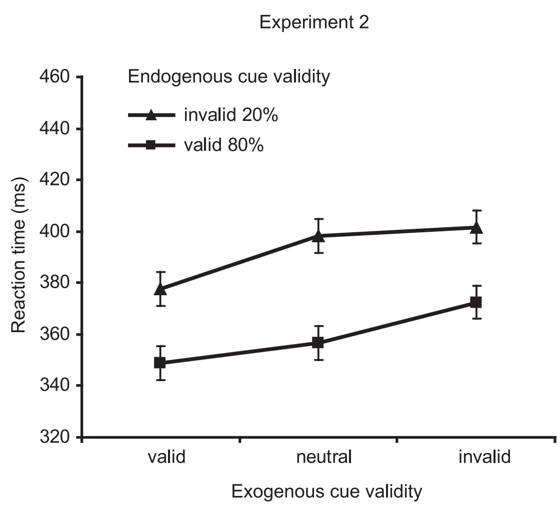

Figure 3. Results of Experiment 2 presented as an average reaction time (ms). The endogenous visual cue conditions (valid 80% and invalid 20%) are plotted as separate lines. The exogenous auditory cue conditions (valid, neutral, and invalid) are plotted along the *x*-axis. The error bars show the .95 confidence intervals for the exogenous cuing main effect.

significant effect. No effect for auditory neutral compared with auditory invalid conditions (402 ms) was found ($t < 1$).

These results suggest that the observed two-way interaction is because the effect of the visual cue is different in the neutral condition than in the conditions in which the auditory cue provided location information. To investigate this notion, we conducted an additional ANOVA with visual cuing (valid and invalid) and auditory cuing (valid and invalid) as within-subject variables. An endogenous visual cuing effect, $F(1, 11) = 30.729$, $MSE = 327.475$, $p < .001$, and an exogenous auditory cuing effect, $F(1, 11) = 50.906$, $MSE = 134.369$, $p < .001$, were found. The two-way interaction between visual and auditory cuing was not significant ($F < 1$), suggesting that both cuing effects occur independently from one another.

The mean error scores for the valid endogenous cuing condition (auditory cue valid, 3.0%; neutral, 3.6%; and invalid 5.2%) and the invalid endogenous cuing condition (auditory cue valid, 6.0%; neutral, 8.3%; and invalid, 8.0%) were also calculated. An ANOVA showed a significant main effect for visual cuing, $F(1, 11) = 5.691$, $MSE = .004$, $p < .05$; no effect for auditory cuing, $F(2, 22) = 2.014$, $MSE = .001$, $p = .157$; and no interaction effect between visual and auditory cuing ($F < 1$).

Discussion

Even though participants focused their visual attention on a location in space before the presentation of the auditory cue, there was still a reliable auditory cuing effect. These effects replicate earlier results of van der Lubbe and Postma (2005) and are in line with the results found by Mazza and colleagues (2007). Furthermore, a strong endogenous visual cuing effect was found, confirming that the participants used this cue to enhance their performance.

When we compare the results for the condition in which the endogenous visual cue was valid with those of Experiment 1, we see that the exogenous cuing effects are similar. When the locations of the auditory cue and target are congruent (valid cue condition), there is a decrease in reaction time relative to the neutral cue condition, and when the locations of auditory cue and target are incongruent (invalid cue), there is an increase relative to the neutral condition. Thus, the data suggest that even when participants are focused on the target location, they still benefit from an additional valid auditory cue. There are also reaction time costs in the invalid auditory cue condition, indicating that the exogenous cue is able to capture attention while attention is focused. In contrast, for conditions in which the endogenous visual cue was invalid, only a valid auditory cue had an effect compared with the neutral cue condition. In other words, when attention is focused on a nontarget location, an additional invalid auditory cue (which is congruent with the invalid visual cue) has no further effect on response times. However, there is in that case a large effect of the valid auditory cue, suggesting that the exogenous cue helps participants to disengage their attention from the wrong location. Thus the results indicate that sounds facilitate the disengagement of attention from a location, which is beneficial when attention is needed somewhere else.

In summary, we find that in both the visual valid and the visual invalid condition, auditory cues from a location opposite to the attentional focus are able to capture attention. In the case of a valid endogenous visual cue, this capture of attention will have a cost, and in the case of an invalid endogenous cue, it will yield a benefit. In the valid visual condition, we observe an extra benefit of the valid auditory cue, indicating that attention was not completely focused by the endogenous cue alone. This finding can possibly be explained by the fact that the endogenous cue was only valid in 80% of the trials. Yantis and Jonides (1990) have shown that the validity of the endogenous cue strongly influences its ability to suppress exogenous cuing. In their study, visual exogenous cuing effects were suppressed only when a 100% valid endogenous cue was used (see, for similar results, Theeuwes, 1991). However, when the endogenous cue had a 75% validity, an exogenous cuing effect was still observed. According to Yantis and Jonides (1990), the uncertainty concerning the validity of the visual cue could have influenced the way in which participants focused their attention. If in the current experiment participants attention was not fully focused because of this uncertainty, the auditory cue could have improved this focus, resulting in better performance (see also Muller & Rabbitt, 1989). In other words, both the observed costs and benefits can be explained in terms of attention not being completely focused before the location was indicated by the cue. To test this hypothesis, we performed a third experiment in which the endogenous visual cue was valid in 100% of the trials. When a 100% valid endogenous cue is able to fully suppress the capture of visual attention by means of an auditory cue, the exogenous cuing effect should disappear, which would be in line with earlier studies showing a suppression of exogenous cuing when attention is focused endogenously (Santangelo, Belardinelli, & Spence, 2007; Theeuwes, 1991; Yantis & Jonides, 1990).

Experiment 3: Endogenous Cue 100% Correct

In Experiment 3, we used a 100% valid visual endogenous cue to check whether the cross-modal cuing effects found in Experiment 2 are caused by the fact that the endogenous cue was invalid in a small proportion of the trials. If top-down processes are indeed able to suppress exogenous cuing, as was shown earlier within the visual modality, we expect that this manipulation will cause all exogenous cuing effects to disappear.

Method

Participants. Twelve new students of the Vrije Universiteit Amsterdam (2 men, 10 women; mean age = 20.8 years, range = 18–28 years) participated in the experiment.

Design. In Experiment 3, the stimuli and method for eye movement registration were identical to those used in Experiment 2, but endogenous visual cues were used that were valid in 100% of the trials. The design in terms of conditions and amount of trials (48 per condition) was identical to that of Experiment 1.

Results

Trials with eye movements (3.4%) and premature (<200 ms) or slow (>1,000 ms) responses (0.4%) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (96.2%) were calculated for each participant for each condition. The mean reaction times for each condition averaged over participants are plotted in Figure 4. The error bars

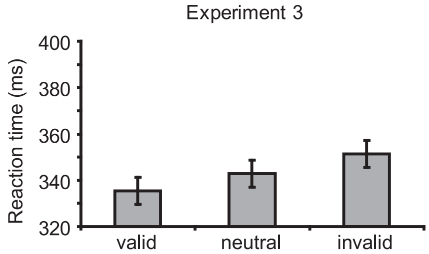

Figure 4. Results of Experiment 3 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cuing main effect.

in this figure represent the .95 confidence interval (5.9 ms) for the exogenous cuing main effect. A within-subjects ANOVA showed that auditory cuing (valid, 335 ms; invalid, 351 ms; and neutral, 343 ms) had a significant effect on the reaction times, $F(2, 22) = 7.836$, $MSE = 97.068$, $p < .005$. Three post hoc pairwise two-tailed t tests between the cuing conditions were conducted. Valid compared with invalid ($p = .008$), neutral compared with valid ($p = .039$), and neutral compared with invalid conditions ($p = .046$) were all significant.

A similar ANOVA applied to the error data (valid, 2.3%; invalid, 4.3%; and neutral, 4.7%) showed a significant cue effect, $F(2, 22) = 4.256$, $MSE < .001$, $p < .05$. Post hoc analysis showed a significant difference only between valid and neutral conditions ($p = .002$), a strong trend for valid compared with invalid conditions ($p = .059$), and no effect for invalid compared with neutral conditions ($t < 1$).

A separate ANOVA for a between-group comparison between Experiment 2 (visual valid) and Experiment 3 showed no significant reaction time differences, $F(1, 22) = 1.257$, $MSE = 2868.708$, $p = .274$.

Discussion

The results of this experiment are similar to those observed in the visual valid condition of Experiment 2. Just as in Experiments 1 and 2, the current results show two effects. Compared with the neutral condition, we see benefits and costs for valid and invalid auditory cues, respectively, which means that the observed cuing effect is not sensitive to the validity of the endogenous visual cue as suggested earlier. These findings are somewhat unexpected because the 100% valid endogenous cue should have allowed a firmer and more focused attention on the target location before the auditory cue was presented (Yantis & Jonides, 1990). Therefore, if cross-modal cuing is based on shifting attention, we would have expected no benefits because shifting attention in response to the auditory cue is not necessary when attention is already focused on the target location. One possibility is that participants did not make full use of the endogenous cue, but this seems unlikely because we already observed a significant effect of endogenous cuing on response times in Experiment 2; also, response times in Experiment 3 are similar to those measured in Experiment 2 for valid endogenous cuing. However, there is still an alternative explanation for the observed costs and benefits. Note that the current paradigm differs with respect to a further aspect in comparison to other studies such as Santangelo and Spence (2007) and Theeuwes (1991).

Theeuwes (1991) showed that cuing effects disappear when there is focused attention in combination with no-onset targets. These no-onset targets in the form of figure-eight premasks (that can turn into letters by removing two of the line segments) were already on screen when the endogenous cue was presented, which allowed participants in response to the endogenous cue to focus their attention tightly on the premasks before the exogenous cue was presented. The same holds for the study by Santangelo and Spence (2007) in which an RSVP stream was presented during the entire trial, also allowing participants to focus their attention. Both the no-onset targets and the RSVP stream could have functioned as placeholders allowing participants to direct their attention to a predefined location in space. It could be the case that the cuing effect observed in the current experiment is the result of not predefining the specific target location. In other words, the observed cuing effect could reflect merely that attention was not entirely focused on one specific location because no placeholders were present. The observed costs and benefits could then simply be explained in terms of shifts (or fine tuning) of attention induced by the exogenous auditory cue. To test this notion, we conducted a fourth experiment in which placeholders were used to indicate target positions.

Experiment 4: Introducing Placeholders

Experiment 4 was similar to Experiment 3 except that placeholders marking the target locations stayed on the screen during the entire trial. These placeholders enabled the participants to accurately focus their attention on the target location. If spatial uncertainty of the target location is indeed the reason why exogenous cuing effects were still observed in the previous experiments, we expect that this manipulation will cause the effects to disappear.

Method

Participants. Twelve new students of the Vrije Universiteit Amsterdam (3 men, 9 women; mean age = 20.8 years, range = 18–25 years) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing.

Stimuli and design. This experiment was similar to Experiment 3, including the 100% valid endogenous cues in the form of arrowheads and the number of trials per condition (48). Additionally, during the entire trial placeholders were displayed on both sides of the screen indicating the possible target locations. The placeholders were thin, gray-lined squares (9.34 cd/m^2) with a width and height of 1.3° that indicated the area in which targets could appear. On each side of the screen two connecting placeholders were shown—one for the targets displayed above the midline of the screen and the other for the targets below the midline. The horizontal center of the squares was separated by an angle of 10.5° from the center of the screen. The placeholders made the task of target discrimination (above or below the vertical middle) easier, because their locations were now predefined. We compensated for this by displaying the targets closer to the vertical middle (on average 0.6° above and below) than in the previous experiments and by varying their location by placing the targets randomly within a range of 0.3° from the center of the placeholder.

Results

Trials containing possible eye movements (5.9%) and premature (<200 ms) or slow (>1,000 ms) responses (0.5%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (93.2%) were calculated for each participant for each condition. Mean reaction times for each condition averaged over participants are plotted in Figure 5. The error bars in this figure represent the .95 confidence interval (7.0 ms) for the exogenous cuing main effect. For the analysis, a within-subjects ANOVA was used, which showed that auditory cuing (valid, 362 ms; invalid, 373 ms; and neutral, 364 ms) had a significant effect on the reaction times, $F(2, 30) = 4.238$, $MSE = 140.244$, $p < .05$. Three post hoc pairwise two-tailed t tests between the cuing conditions were conducted. Valid compared with invalid ($p = .016$) and neutral compared with invalid conditions ($p = .020$) were both significant. Neutral compared with valid conditions ($t < 1$) was not significant. A similar ANOVA applied to the error data (valid, 7.1%; invalid, 6.5%; and neutral, 6.8%) showed no effect ($F < 1$).

Discussion

The results show that the exogenous cuing effect remains but that it is now solely based on costs: Valid cues yield no improvement relative to neutral cues, whereas invalid cues still result in slower reaction times. The results suggest that the effects for valid auditory cues, observed in Experiments 2 and 3, were indeed due to the fact that these cues caused an improved and/or narrowed attentional focus on the location of the impending target. We should also note that the use of placeholders allowed for a smaller vertical angle between the up and down target locations on the left and right side of the screen, which also allowed participants to use a smaller attentional focus in comparison to the previous experiments. All together, the results indicate that exogenous auditory cues can still capture attention, even when visual attention is fully focused and the target location is predefined. Apparently, the auditory cue presented at the nontarget location causes attention to shift away from the target location, causing a small but reliable reaction time cost.

So far, the results of Experiments 1–4 show how exogenous cuing benefits disappear when attention is focused on a valid and predefined target location. For these experiments, we used cross-modal auditory cues that were compared with a neutral condition. It is important to compare these results with results of visual cues

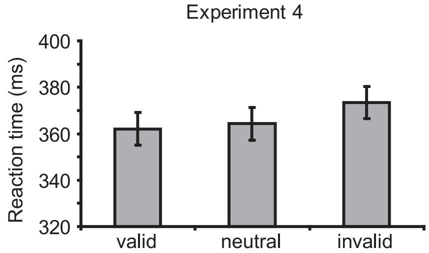

Figure 5. Results of Experiment 4 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cuing main effect.

obtained in identical conditions, because previous results on visual cuing are inconclusive. Recall that Mazza and colleagues (2007) showed no cuing effect for unimodal cues when target location is blocked, but they did show cross-modal cuing effects, Santangelo and Spence (2007) found no unimodal and no cross-modal cuing effects on peripheral targets when attention is focused to the center of the screen, and van der Lubbe and Postma (2005) found both unimodal and cross-modal cuing effects when visual attention is focused on the correct target location.

Experiment 5: Visual Exogenous Cues

In this final experiment, we wanted to determine whether the results shown in Experiment 4 would change when a visual instead of an auditory exogenous cue was used. In other words, the question is whether knowing at which location a target will appear influences unimodal exogenous cuing in the same way as cross-modal exogenous cuing. To test this question, we used the same paradigm as in Experiment 4, but we made the following changes: Endogenous cuing was either 100% valid or absent (a question mark was displayed instead of an arrow), and both visual and auditory exogenous cues were presented. These factors were tested in different blocks in a within-subjects design. The conditions without endogenous cuing were included so that we could verify that exogenous cuing also occurs also in the absence of an endogenous cue. We used a modified set-up that allowed us to exactly align the auditory cue with the visual targets, to prevent spatial disparities. We did this by displaying the visual stimuli on an acoustically transparent screen and by placing the loudspeakers at the exact target locations (10.5° left and right from the middle).

Method

Participants. Sixteen new students of the Vrije Universiteit Amsterdam (2 men, 14 women; mean age = 20.2 years, range = 18–25 years) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing.

Apparatus. In the experiment, the visual stimuli were presented on a sound transparent (microperforated) screen by means of a projector (Theme Scene HD70, 60 Hz; Optoma Europe, Hertfordshire, UK). Participants were seated in a dimly lit room approximately 150 cm from the screen. All visual stimuli were rescaled so their retinal images were of identical size as in the previous experiments.

Stimuli and design. Compared with Experiment 4, two within-subject factors were added to the experiment's design. First, we presented either a 100% valid endogenous cue (an arrowhead pointing left or right) or a neutral cue in the form of a question mark (?) of similar size. Second, the exogenous cue (which could be valid, neutral, or invalid) was auditory or visual. This visual cue was a thinly lined dark gray circle with a diameter of 3.8° that was flashed for 100 ms at the target location (10.5° left or right from the middle). In the neutral condition, the circles were flashed simultaneously at both sides. The onset of the exogenous cues occurred 200 ms before the onset of the target. The two additional factors were tested in four conditions that were presented in a blocked fashion in the form of four subexperiments. The following combinations were presented:

1. Auditory exogenous cue (valid, neutral, invalid)—no endogenous cue.

2. Auditory exogenous cue—endogenous cue.

3. Visual exogenous cue (valid, neutral, invalid)—no endogenous cue.

4. Visual exogenous cue—endogenous cue.

Each subexperiment contained five blocks containing 40 trials for each exogenous cuing condition (a total of 24 trials per block). The order in which the four subexperiments were presented to each participant was balanced by means of a Latin square. The session started with one practice block of 24 trials that was identical to the first bock of the participants' first subexperiment.

Results

Trials containing possible eye movements (4.3%) and premature (<200 ms) or slow (>1,000 ms) responses (0.3%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (94.6%) were calculated for each participant for each condition. Mean reaction times per condition are shown in Figure 6. The error bars in this figure represent the .95 confidence interval (9.8 ms) for the exogenous cuing main effect. For the analysis, a within-subjects ANOVA was conducted containing the factors exogenous cue modality (visual, auditory), exogenous cue validity (valid, neutral, invalid), and endogenous cue presence (cue, no cue). This ANOVA showed a main effect of exogenous cue modality, $F(1, 15) = 90.395$, $MSE = 657.465$, $p < .001$, indicating an overall faster reaction time on targets preceded by an auditory cue than by a visual cue. There was also a main effect of exogenous cue validity, $F(2, 30) = 81.052$, $MSE = 370.603$, $p < .001$, and a main effect of the presence of the endogenous cue, $F(1, 15) = 7.941$, $MSE = 1074.622$, $p < .05$, indicating an overall faster performance when the target was preceded by a valid endogenous cue relative to a no cue condition. Additionally, there was a two-way interaction between exogenous cue modality and exogenous cue validity, $F(2, 30) = 17.023$, $MSE = 331.302$, $p < .001$, indicating an overall stronger exogenous cuing effect for visual cues. Also, a two-way interaction was found between endogenous cue presence and exogenous cue validity, $F(2, 30) = 4.463$, $MSE = 196.075$, $p < .05$, suggesting a reduced exogenous cuing effect when a valid endogenous cue is present, which is in line with the results from Experiment 4 that showed a reduced cuing effect in comparison to Experiment 1. This reduction is primarily based on the disappearance of benefits when the valid target location is endogenously cued. No interaction between exogenous cue modality and endogenous cue presence was observed ($F < 1$). However, more important, there was no three-way interaction, $F(2, 30) = 1.268$, $MSE = 233.245$, $p = .296$, suggesting that there is no difference in the way that the endogenous cue interacts with visual or auditory exogenous cues.

A similar ANOVA conducted on the error data showed no significant effects but only a trend, $F(2, 30) = 2.964$, $MSE = .001$, $p = .067$, for exogenous cue validity (valid, 4.9%; neutral, 5.6%; and invalid, 6.1%). These results indicate that there was no speed–accuracy trade-off.

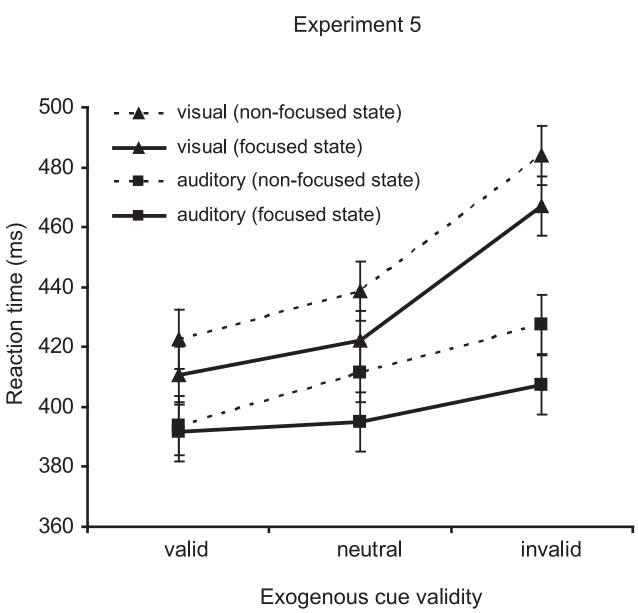

Figure 6. Results of Experiment 5 presented as an average reaction time (ms) for each condition. The four combinations for the factors exogenous cue modality (visual and auditory) and endogenous cue presence (present, focused state, and not present, nonfocused state) are plotted as separate lines. The endogenous cue validity (valid, neutral, and invalid) is plotted on the *x*-axis. The error bars show the .95 confidence intervals for the exogenous cuing main effect.

Discussion

The results of this experiment show similar effects for both visual and auditory cues. The results for auditory cross-modal exogenous cues replicated those of Experiment 4 by showing a reduced exogenous cuing effect when visual attention is focused on the correct target location. In addition, the current results show that the spatial disparity between auditory cues and visual targets when presented in all previous experiments had no noticeable influence on the main cuing effect: The new set-up used in this experiment, which allowed us to present auditory cues and visual target at the same location, showed the same cross-modal cuing effects. A similar interaction between endogenous cue presence and exogenous cue validity is present in the visual unimodal exogenous cue condition. Overall these results indicate similar attentional processes involved in unimodal and cross-modal cuing and are in line with the results shown by van der Lubbe and Postma (2005).

General Discussion

This study was conducted for three reasons: First, to investigate whether an endogenous visual cue can suppress exogenous cross-modal cuing. In other words, is exogenous cross-modal cuing an automatic process, or are top-down processes able to influence exogenous cross-modal cuing? As mentioned, the results from previous studies (Mazza et al., 2007; Santangelo et al., 2007; van der Lubbe & Postma, 2005) are inconclusive on this issue. Second, we wanted to examine how both costs and benefits contribute to the cross-modal cuing effect. To test this issue, we introduced a neutral exogenous cue that was spatially uninformative but still provided the same temporal information as the valid and invalid cues. Faster responses in valid conditions compared with in neutral conditions would indicate benefits, and slower responses in invalid conditions compared with in neutral conditions would reflect costs (Posner, 1980). A neutral baseline condition has not been used in earlier studies investigating endogenous and exogenous cross-modal cuing. Third, we wanted to test whether an endogenous cue would influence cross-modal and unimodal exogenous cuing in the same way.

The first experiment not only replicated the findings of Spence and Driver (1997) but also showed that cross-modal cuing generates both reaction time costs and benefits. Similar to notions based on unimodal cuing (Posner, 1980), these results suggest that cuing effects reflect shifts of attention. In the following experiments, an endogenous visual cue (a centrally presented arrowhead) was presented before the presentation of the exogenous auditory cue. The results from Experiments 2 and 3 showed that when attention was in a focused state, there was still a cuing effect observed consisting of both costs and benefits. Comparison of the results of Experiments 2 and 3 also indicates that the cuing effect is not sensitive to the validity of the endogenous visual cue. However, when attention can be focused on a predefined target location by means of placeholders as in our Experiment 4, only costs are observed and no benefits. For the auditory cuing conditions, the results of Experiment 5 replicate those of Experiments 1 and 4. In addition, the results show that visual exogenous cues, at least with the visual task that we have used, in essence have the same effect as auditory cues.

As to the first question of this study, whether an endogenous cue can suppress the exogenous capture of attention by an exogenous cross-modal event, the results from Experiments 2–5 provide compelling evidence that the answer should be “no.” These results show that when attention is in a focused state by means of a centrally presented arrowhead (Experiment 2) that is 100% valid (Experiment 3) and pointing to a predefined target location (Experiments 4 and 5), there is still an effect of the presence of an exogenous auditory and visual (Experiment 5) event. These results extend those of van der Lubbe and Postma (2005) and Mazza and colleagues (2007) by showing benefits and costs in the form of attentional capture for both valid and invalid cues, respectively, and how this capture strongly depends on the attentional focus before the presentation of these cues.

The second question, whether cross-modal cuing effects are based on costs, benefits, or both, could indeed be answered by introducing conditions with a neutral cue, and it appeared that the costs and benefits depend strongly on how spatial attention is focused. As stated earlier, these results are in line with the general view that cuing effects are based on shifts of attention. Remarkably, both Experiments 2 and 3 showed benefits when visual attention was focused on the valid target location by means of an endogenous visual cue presented before the onset of the exogenous auditory cue. When benefits are indeed based on attentional shifts, this result should not be expected: When attention is already focused on the correct target location, additional spatial information should not result in an extra performance improvement. If anything, one would have expected additional costs because the auditory cue was presented at an eccentricity that was larger than the eccentricity at which the target was presented. In other words, presenting auditory cues from loudspeakers positioned next to a monitor could have resulted in attention being drawn to the loudspeaker location rather than to the target location. Apparently this was not the case even though this was a concern when designing the display set-up for this study. The results of Experiment 4 show that performance improvements due to valid cues disappear when the target locations are predefined by means of placeholders, which indicates that an exogenous auditory cue is able to influence only attentional focus when there is uncertainty with respect to target position. Note that attentional capture—the cost associated with an invalid exogenous cue—is affected neither by the validity of the endogenous cue nor by the spatial uncertainty of the target location.

The third question, whether an endogenous visual cue would influence cross-modal and unimodal exogenous cuing the same way, was answered by Experiment 5. These results show a reduced exogenous cuing effect in both unimodal and cross-modal conditions, when attention is in a focused state. These results are in line with those of van der Lubbe and Postma (2005). Although Santangelo and Spence (2007) also showed that focused attention influences cross-modal and unimodal exogenous cuing similarly, their results are completely opposite. Note that Mazza and colleagues (2007) are the only ones who showed within one study opposite results for unimodal and cross-modal cuing. When they provided endogenous information by blocking target side, the results showed no unimodal cuing effect. However, the same experiment did show a cuing effect for the cross-modal condition. However, it must be noted that by blocking trials, they have potentially introduced a confound that complicates the interpretation of their findings. As explained by Santangelo and colleagues (2007), a drawback of blocking target side is that “the presentation of a target on one side on one trial may lead to an exogenous shift of attention toward that side, thus potentially facilitating performance when the target on the next trial also happens to be presented from the same side” (p. 138). Thus, their results might reflect only a differential influence of this trial-to-trial within-modality cuing on the actual cuing effect that was under study.

As already mentioned in the introduction, there is no simple way to explain why some studies found suppression of exogenous attention and others did not. Still, there are some important differences that could play a role. Most studies adopted the task of Spence and Driver (1997) that uses dots as targets, which appeared in the form of onsets. This method is in contrast to the no-onset letter stimuli used earlier by Theeuwes (1991). It may be that it is easier for participants to ignore onset cues when the targets are no-onsets instead of onsets, which is consistent with the notion of contingent capture (Folk, Remington, & Johnston, 1992), which states that exogenous capture of attention by, for instance, an irrelevant cue depends on whether the cue shares a relevant feature with the target. Folk and his colleagues (1992) showed that onset cues affect onset targets but not targets that are characterized by a color change. In other words, when onset targets are used, onsets become a relevant stimulus feature. As a result, the onset of the cue may automatically draw attentional resources to the cue location, which could explain the cuing effect during focused attention.

Letter stimuli were also used in Santangelo and Spence’s (2007) task, but this time in the form of a central RSVP stream on which attention should be focused. This RSVP stream presented letters at a high rate, and knowing that a possible target could appear in this stream kept participants endogenously focused to it. However, an RSVP stream also tends to generate a high perceptual load, which in turn might drain attentional resources required for the processing of the exogenous cues. A follow-up study by Santangelo, Finoia, Raffone, Belardinelli, and Spence (2008) using a central morphing shape (instead of an RSVP stream) to manipulate purely perceptual load confirms this idea by again showing suppression of exogenous visual cuing. Taken together, both endogenous attention and perceptual load could explain the suppression of the unimodal and cross-modal exogenous cuing effect as shown by Santangelo and Spence (2007). In a recent review, Santangelo and Spence (2008) discussed whether unimodal cuing and cross-modal cuing are automatic processes. They evaluate exogenous cuing by means of the intentionality and load-insensitivity criteria (Jonides, 1981; Posner, 1978; Yantis & Jonides, 1990), stating that voluntary control and perceptual load should not interfere with a process in order for it to be automatic. Santangelo and Spence claimed that when an RSVP stream is used to focus attention it is hard to distinguish between possible voluntary endogenous effects of the task (find the target in the stream) and perceptual load effects evoked by the information presented in the RSVP stream. On the basis of their findings, they concluded that the capability of abrupt onsets to capture spatial attention depends on how much attentional resources are available. If one's resources are fully engaged by means of a high perceptual load task such as an RSVP stream, there will probably be no attentional capture effects. By contrast, if an endogenous cue is used to voluntarily focus attention, it is likely that there will be enough resources left to process peripheral onsets that are able to capture attention. Therefore, our finding that attentional capture by means of exogenous cues cannot be suppressed when attention is focused in a pure endogenous fashion is not necessarily inconsistent with the views proposed by Santangelo and Spence (2008).

In conclusion, the results from these current experiments make clear that cross-modal and unimodal exogenous cuing of a visual target location cannot be suppressed by endogenously focusing visual attention. Even when visual attention is fully focused to a predefined target location, an auditory or visual cue coming from the opposite direction is still able to capture visual attention. When visual attention is not focused before the presentation of the auditory cue, both costs and benefits are shown.

References

Blauert, J. (1997). *Spatial hearing: Psychophysics of human sound localization.* Cambridge, MA: MIT Press.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica, 50,* 253–290.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science, 12,* 360–365.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movements. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin and Review, 1,* 476–490.

Mazza, V., Turatto, M., Rossi, M., & Umilta, C. (2007). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia, 45,* 514–522.

McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science, 11,* 167–171.

Mondor, T. A., & Amirault, K. J. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 745–755.

Muller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual-attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330.

Peelen, M. V., Heslenfeld, D. J., & Theeuwes, J. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *NeuroImage, 22,* 822–830.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Psychology Software Tools. (1996–2003). E-Studio (1.1.4.15), E-Prime 1.1 (SP3) (1.1.4.4). Pittsburgh, PA: Author.

Santangelo, V., Belardinelli, M. O., & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 137–148.

Santangelo, V., Finoia, P., Raffone, A., Belardinelli, M. O., & Spence, C. (2008). Perceptual load affects exogenous spatial orienting while working memory load does not. *Experimental Brain Research, 184,* 371–382.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1311–1321.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition, 17,* 989–1015.

Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 555–574.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1005–1030.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59,* 1–22.

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 277–320). Oxford, England: Oxford University Press.

Spence, C., & McGlone, F. P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research, 141,* 324–330.

Talsma, D., & Woldorff, M. G. (2005a). Methods for the estimation and removal of artifacts and overlap in ERP data. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 115–148). Cambridge, MA: MIT Press.

Talsma, D., & Woldorff, M. G. (2005b). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17,* 1098–1114.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception and Psychophysics, 49,* 83–90.

van der Lubbe, R. H. J., & Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental Brain Research, 164,* 464–471.

Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception and Psychophysics, 62,* 1258–1264.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual-search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121–134.

Received October 19, 2007
Revision received August 11, 2008
Accepted August 13, 2008 ■